Figure S1

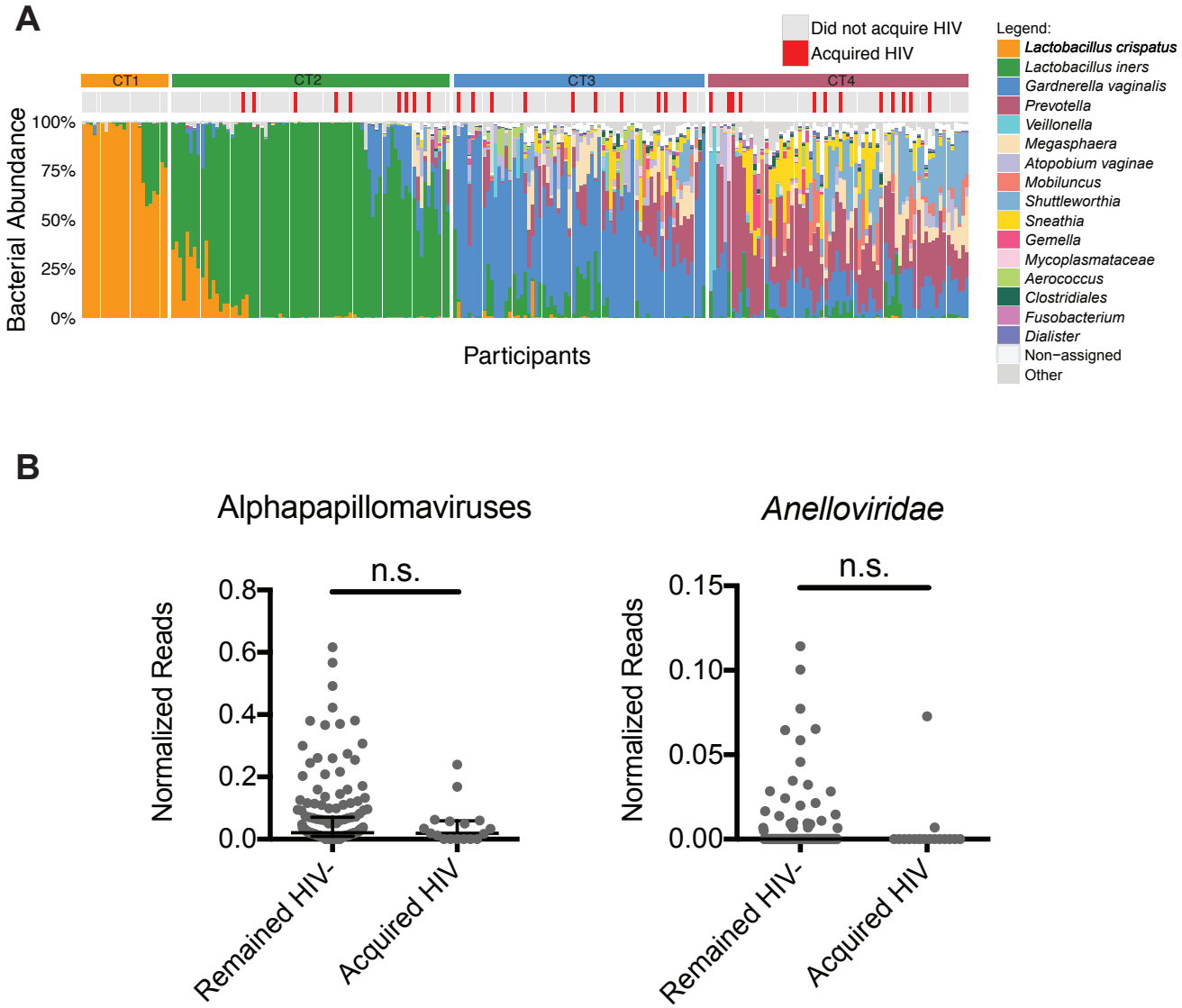


Figure S1. FGT bacteria and viruses in participants who subsequently acquired HIV. Related to Figure 2. (A) Stacked bar plot showing the bacterial composition of the cervicovaginal microbiome of 236 HIV-uninfected women. Participants who subsequently acquired HIV (n=31) are indicated. **(B)** Alphapapillomaviruses and *Anelloviridae* are not associated with HIV acquisition. Normalized reads of alphapapillomaviruses and *Anelloviridae*, grouped by women who remained HIV-uninfected (n=163) and those who became infected with HIV (n=17). Lines indicate the median and the interquartile range (IQR). Groups were compared using Mann-Whitney test.

Figure S2

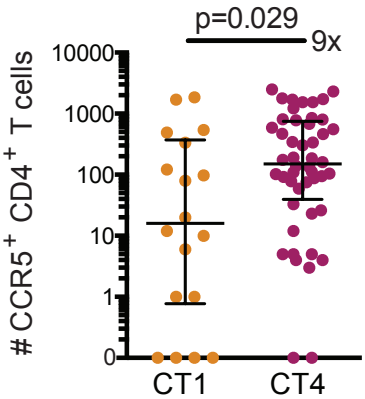
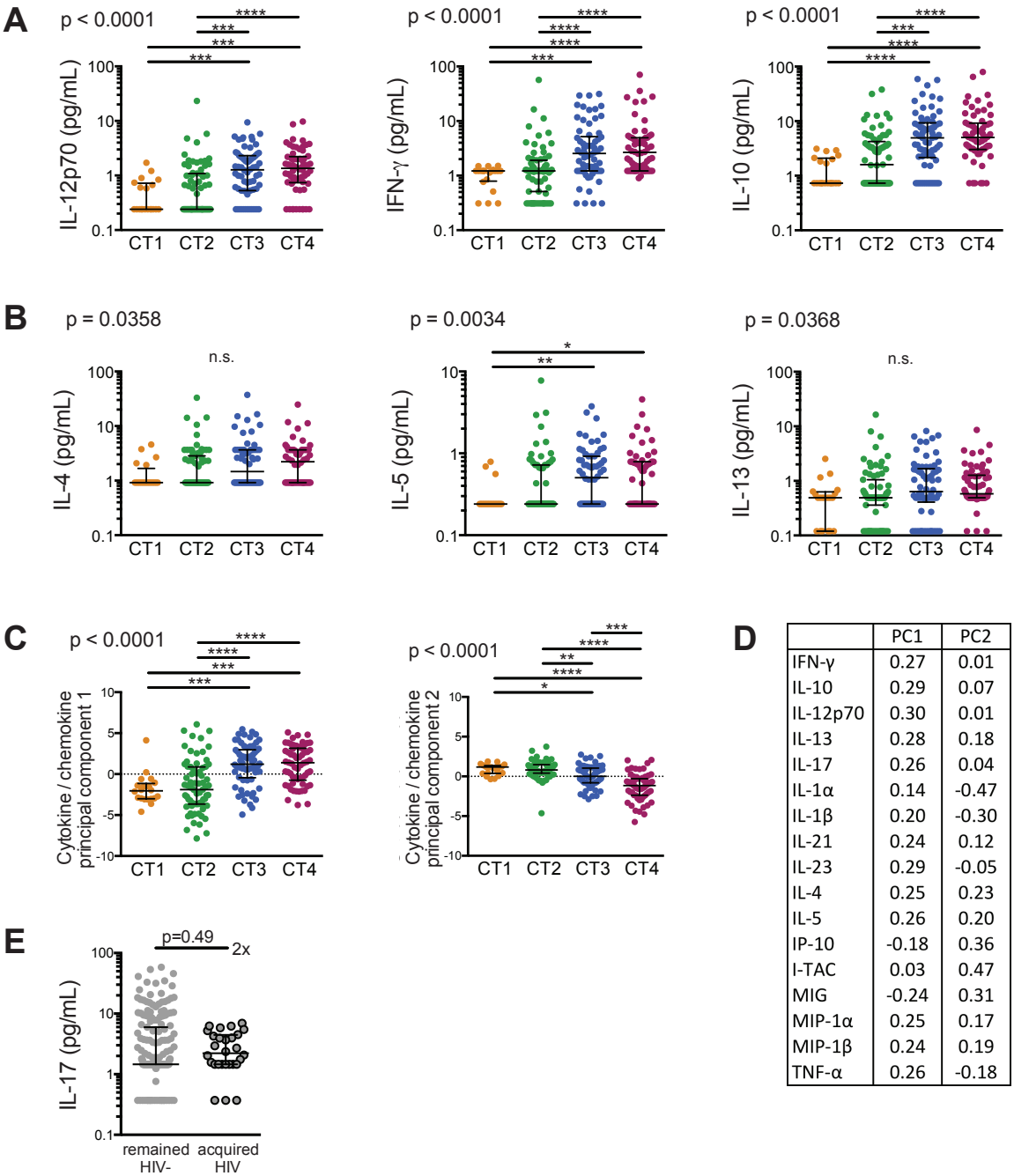


Figure S2. Increased mucosal CCR5⁺ CD4⁺ T cell numbers in participants with CT4 bacterial communities compared to individuals with CT1 communities. Related to Figure 3.

Flow cytometry analysis of cells isolated from cytobrushes from 66 individuals (n (CT1) = 18; n (CT4) = 48). Lines in the plots indicate the median and the interquartile range (IQR) of the dataset. Groups were compared by Mann-Whitney test.

Figure S3



D

	PC1	PC2
IFN-γ	0.27	0.01
IL-10	0.29	0.07
IL-12p70	0.30	0.01
IL-13	0.28	0.18
IL-17	0.26	0.04
IL-1α	0.14	-0.47
IL-1β	0.20	-0.30
IL-21	0.24	0.12
IL-23	0.29	-0.05
IL-4	0.25	0.23
IL-5	0.26	0.20
IP-10	-0.18	0.36
I-TAC	0.03	0.47
MIG	-0.24	0.31
MIP-1α	0.25	0.17
MIP-1β	0.24	0.19
TNF-α	0.26	-0.18

Figure S3. Additional cytokines measured in CVL fluid, and PCA loadings. Related to Figure 3.

(A, B) Chemokine and cytokine concentrations in CVL samples from 219 individuals, grouped by CT. **(C)** Cytokine principal component (PC) 1 and PC2 values for 219 subjects, grouped by CT. Groups in (**A**)-(**C**) were compared using Kruskal-Wallis test with Dunn's post-hoc analyses. **(D)** Cytokine loadings on PC1 (51.3% of variation) and PC2 (14.6% of variation). **(E)** IL-17 concentrations in the CVL fluid of participants who remained HIV negative (n=28) and those who acquired HIV (n=191) (Mann-Whitney test). Lines indicate the median and the interquartile range (IQR) of each dataset.

Figure S4

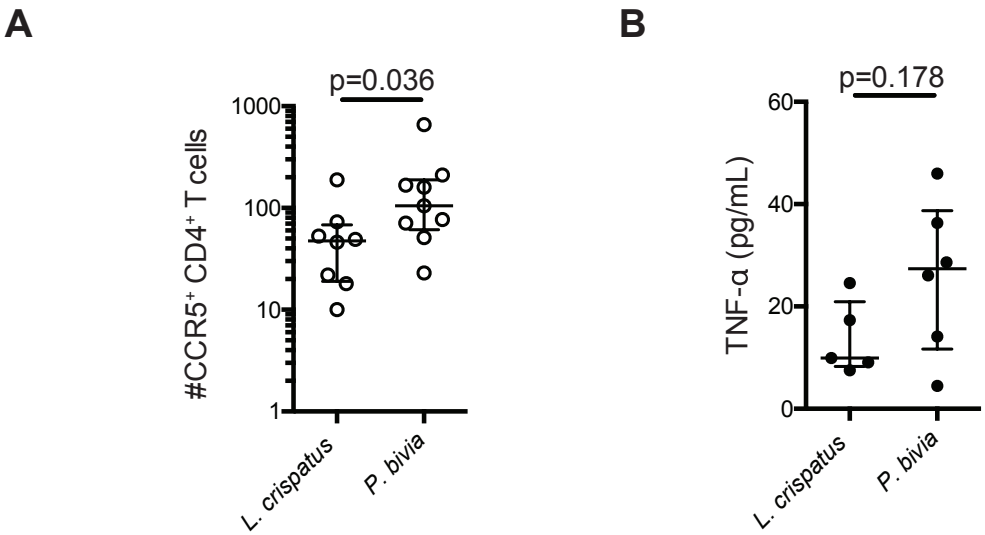


Figure S4. Genital mucosal CCR5⁺ CD4⁺ T cells and TNF-α concentrations in germ-free mice intravaginally inoculated with *L. crispatus* and *P. bivia.* Related to Figure 4.

(A) Flow cytometry analysis of cells isolated from the FGT. Pooled results from 4 independent experiments.

(B) TNF-α concentrations measured in CVL fluid. Pooled results from 3 independent experiments.

Lines indicate the median and the interquartile range (IQR) of each dataset. Groups were compared using Mann-Whitney test.

Table S1

	All n=236	CT1 n=23	CT2 n=75	CT3 n=68	CT4 n=70	
Age of participant at enrollment (Median, IQR)	21.4 [20, 22.4]	20.7 [19.3, 22.1]	21.5 [20.1, 22.4]	21.5 [20.4, 22.7]	21.4 [20.2, 22.6]	p = 0.2352[e]
Age difference between participant and current sexual partner (Median, IQR)[f]	2.9 [1, 5.1]	2.5 [1.2, 3.7]	2.9 [0.6, 5]	3 [1.8, 5.9]	3 [1, 5.1]	p = 0.6630[e]
STIs:	% (n/t)	% (n/t)	% (n/t)	% (n/t)	% (n/t)	
Neisseria gonorrhoeae	4% (10/236)	9% (2/23)	3% (2/75)	4% (3/68)	4% (3/70)	p = 0.5889[a]
Chlamydia trachomatis	16% (37/236)	0% (0/23)	13% (10/75)	16% (11/68)	23% (16/70)	p = 0.0446[a]
Mycoplasma genitalium	6% (15/236)	4% (1/23)	8% (6/75)	4% (3/68)	7% (5/70)	p = 0.8562[a]
Trichomonas vaginalis	8% (18/236)	4% (1/23)	7% (5/75)	7% (5/68)	10% (7/70)	p = 0.8735[a]
Herpes simplex virus 2	2% (5/217)	0% (0/23)	0% (0/69)	3% (2/63)	5% (3/62)	p = 0.2541[a]
Herpes simplex virus 1	0% (0/195)	0% (0/19)	0% (0/64)	0% (0/59)	0% (0/53)	p = 1[a]
Experiencing STI symptoms at time of exam: [c]						
No	92% (217/236)	100% (23/23)	91% (68/75)	97% (66/68)	86% (60/70)	p = 0.1530[b]
Yes	6% (14/236)	0% (0/23)	5% (4/75)	3% (2/68)	11% (8/70)	
Not available	2% (5/236)	0% (0/23)	4% (3/75)	0% (0/68)	3% (2/70)	
Contraceptives:						
No family planning	52% (121/232)	61% (14/23)	48% (35/73)	45% (30/67)	61% (42/69)	p = 0.4540[b]
DMPA/Nuristerate/Implanon	44% (102/232)	39% (9/23)	48% (35/73)	52% (35/67)	33% (23/69)	
Oral contraception	3% (6/232)	0% (0/23)	3% (2/73)	3% (2/67)	3% (2/69)	
Other[d]	1% (3/232)	0% (0/23)	1% (1/73)	0% (0/67)	3% (2/69)	
Condoms in 30 days:						
Always	21% (48/236)	26% (6/23)	13% (10/75)	18% (12/68)	30% (21/70)	p = 0.0508[b]
Sometimes	34% (80/236)	35% (8/23)	41% (31/75)	34% (23/68)	26% (18/70)	
Never	20% (47/236)	13% (3/23)	27% (20/75)	25% (17/68)	11% (8/70)	
Not available	25% (58/236)	26% (6/23)	19% (14/75)	24% (16/68)	33% (23/70)	
Drying agent usage:						
Always	1% (2/236)	0% (0/23)	1% (1/75)	1% (1/68)	0% (0/70)	p = 0.4684[b]
Sometimes	10% (24/236)	13% (3/23)	9% (7/75)	15% (10/68)	6% (4/70)	
Never	74% (174/236)	65% (15/23)	79% (59/75)	66% (45/68)	79% (55/70)	
Not available	15% (36/236)	22% (5/23)	11% (8/75)	18% (12/68)	16% (11/70)	
# of sex episodes in past 30 days:						
One or more	74% (173/236)	70% (16/23)	80% (60/75)	76% (52/68)	67% (47/70)	p = 0.2649[b]
None	24% (56/236)	26% (6/23)	17% (13/75)	24% (16/68)	31% (22/70)	
Not available	2% (4/236)	4% (1/23)	3% (2/75)	0% (0/68)	1% (1/70)	
# of anal sex episodes in past 30 days:						
One or more	3% (6/236)	0% (0/23)	4% (3/75)	1% (1/68)	3% (2/70)	p = 0.6078[b]
None	72% (170/236)	74% (17/23)	76% (57/75)	75% (51/68)	64% (45/70)	
Not available	25% (60/236)	26% (6/23)	20% (15/75)	24% (16/68)	33% (23/75)	
# of regular sex partners in last 30 days:						
One	70% (165/236)	65% (15/23)	77% (58/75)	71% (48/68)	63% (44/70)	p = 0.4258[b]
More than one	0% (0/236)	0% (0/23)	0% (0/75)	0% (0/68)	0% (0/70)	
None	4% (9/236)	4% (1/23)	1% (1/75)	6% (4/68)	4% (3/70)	
Not available	26% (62/236)	30% (7/23)	21% (16/75)	24% (16/68)	33% (23/70)	
# of casual sex partners in last 30 days:						
One	4% (10/236)	0% (0/23)	4% (3/75)	4% (3/68)	6% (4/70)	p = 0.6433[b]
More than one	0% (0/236)	0% (0/23)	0% (0/75)	0% (0/68)	0% (0/70)	
None	70% (164/236)	70% (16/23)	76% (56/75)	72% (49/68)	63% (43/70)	
Not available	27% (62/236)	30% (7/23)	22% (16/75)	24% (16/68)	34% (23/70)	

a. Fisher's exact test; comparing the number of participants with each STI to those without that STI.
b. Fisher's exact test, comparing the number of participants in all the groups lists under the subheading.
c. Symptoms defined as itching, pain, burning, sores, foul smelling discharge, blood, and pain during sex.
d. Includes women reporting use of Loop/IUD or that underwent tubal ligation
e. Kruskal-Wallis test, shown with median and interquartile range.
f. Participant's age is subtracted from the age of the current partner

Table S1. Active sexually transmitted infections (STIs), contraceptive, condom and drying agent usage, and sexual behavior according to cervicotype (CT) for 236 HIV-uninfected individuals. Related to Figure 2.

Supplemental Experimental Procedures

Study cohort

The individuals analyzed in this study were enrolled in the Females Rising through Education, Support, and Health (FRESH) study, a prospective observational study conducted near Durban, South Africa, a region with high HIV prevalence. Participants were followed between Nov 26, 2012 and Sep 15, 2016 for a median of 336 days (IQR: 178.5 to 347 days) and a total of 198.2 person-years. The study protocol was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal and the Massachusetts General Hospital Institutional Review Board (2012P001812/MGH). Informed consent was obtained following the explanation of the nature and possible consequences of the study. Participants received intensive HIV prevention counselling, and condoms (both female and male) were provided at the study site.

Eligibility criteria

To be eligible for the study, participants had to be female, 18-23 years old, sexually active, HIV-negative, and able to understand the provided information and consent forms. They further had to be willing to adhere to study requirements, to have HIV tests performed twice per week and to have samples stored. Criteria for exclusion were pregnancy, anemia or enrolment in another study.

Clinical procedures

Twice per week, participants attended classes focused on personal empowerment, job skills training, and HIV prevention, and underwent finger prick blood draw for HIV RNA viral load testing by PCR. Every 3 months, participants had a peripheral blood draw and a pelvic exam (not performed during menstruation) that involved the collection of

ectocervical and midvaginal swabs (Catch-All, Epicentre), a cervicovaginal lavage (CVL), and an endocervical cytobrush as described previously (Anahtar et al., 2015). All sample collection was performed by a single nurse throughout the entire study. The participants further completed a detailed HIV risk questionnaire that was administered by a counselor and addressed the participant's sexual behavior, STI history, antibiotic usage, diet, and family planning.

Outcomes

Time to HIV-1 infection was defined as the time from enrolment to the first confirmed positive HIV RNA PCR result, or to the last day of viral load assessment for individuals that remained HIV-negative. Assessment of HIV status was made at study entry and twice per week during follow-up.

Selection of samples for analysis

At the time of data analysis, pre-infection mucosal samples from 31 women who became HIV-infected, and samples from 205 women who remained uninfected were available. The estimated median time to infection after the most recent pelvic exam was 35 days (IQR: 25 to 68.5 days), assuming that infection occurred on average 7 days prior to detection of HIV in the blood.

16S rRNA gene sequencing was performed at a single time point for all 236 participants, using the most recent pre-infection ectocervical swab for women who acquired HIV, and the first collected ectocervical swab for women who remained HIV-negative. We chose the latter based on the observed high stability of the vaginal microbiota in FRESH participants over time (Anahtar et al., 2015), and based on the availability of a larger number of cytobrush samples that had flow cytometry performed at this time point. All

further analyses (virome sequencing, Luminex, flow cytometry (HIV target cell analysis), STI and behavioral data) were conducted for the same time point as the 16S rRNA analysis, provided the availability of a matched sample. If no matched sample was available, the individual was excluded from the analysis. We performed Luminex analysis of 219 available paired CVL samples, including 28 pre-infection samples, virome analysis of 180 paired samples, including 17 pre-infection samples, and target cell frequency analysis of 169 paired samples, including 12 pre-infection samples.

Nucleic acid extraction from cervical swabs, PCR amplification and sequencing of the V4 region of the bacterial 16S rRNA gene

This method was described in detail previously (Anahtar et al., 2016; Anahtar et al., 2015).

Analysis of 16S sequencing results was performed using MacQIIME (Caporaso et al., 2010). Fastq files were quality filtered (excluding sequences with a Phred quality score < 30) and de-multiplexed using the split_libraries.py command. The resulting fasta files were combined, and operational taxonomic units (OTUs) were assigned using open-reference picking (97% identity, Greengenes v.13.8) with default parameters except for 0.1% subsampling. Following OTU picking, a median sequence count of 19,426 was obtained, with a minimum count of 5,175 per sample. For display of the bacterial community structure, taxa were summarized at the lowest taxonomic rank resolved by OTU picking, that is at the species level or higher. The core_diversity_analyses.py workflow was used to calculate Faith's Phylogenetic Diversity (diversity within a sample), and the Principal Coordinates Analysis (PCoA) plots in EMPeror (Vazquez-Baeza et al., 2013), using the weighted UniFrac method (Lozupone and Knight, 2005) with a rarefaction cutoff of 5,000 reads per sample.

Definition of cervicotypes

Cervicotypes (CTs) were defined as previously described (Anahtar et al., 2015). Briefly, samples with the relative majority of sequences assigned by QIIME open-reference OTU picking to the genus *Lactobacillus* (but not *Lactobacillus iners*) were defined as CT1. Shotgun sequencing and oligotyping analyses demonstrated that *Lactobacillus* species other than *L. iners* were mostly composed of *Lactobacillus crispatus* (Anahtar et al., 2015). *Lactobacillus iners* was the predominant species in samples assigned to CT2, and *Gardnerella vaginalis* in samples assigned to CT3. CT4 samples have a dominant bacterial taxon other than *Lactobacillus*, *L. iners* or *G. vaginalis*.

Virome Sequencing

For each sample, 300 μl of PBS was added to 200 μl of frozen CVL fluid. The CVL fluid was passed through a 0.45-μm-pore-size membrane, and DNase and Lysozyme treated to enrich for virus-like particles. Total RNA plus DNA was extracted using the Qiagen DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's recommendation. Purified nucleic acid was reverse transcribed and PCR amplified using barcoded primers consisting of a base-balanced 16-nucleotide-specific sequence upstream of a random 15-mer as previously described (Finkbeiner et al., 2009) and used for NEBNext library construction (New England BioLabs). Libraries were multiplexed (12 samples per flow cell) on an Illumina MiSeq instrument (Washington University Center for Genome Sciences) and sequenced using the paired-end 2 × 250 protocol.

Detection of viral sequences was done using VirusSeeker, a custom bioinformatics pipeline designed to detect sequences sharing nucleotide and protein level sequence

similarity to known viruses (Handley et al., 2016; Monaco et al., 2016). In brief, potential unique viral reads were queried against the NCBI nt/nr databases, and only reads matching exclusively to viral sequences were kept for further analysis. All sequences aligning to viruses were further classified into viral families based on the NCBI taxonomic identity of the best hit. Absolute read counts were normalized by dividing individual taxon sequence counts by the total assigned sequence count in a sample. Richness and diversity were calculated using the diversity function of the *vegan* package (Oksanen, 2016).

Sexually transmitted infection (STI) testing

HIV-1 infection was assessed with a rapid screening HIV RNA PCR assay. Infection with *Chlamydia trachomatis, Neisseria gonorrhoea, Trichomonas vaginalis*, HSV-1 and HSV-2 was tested using a posterior fornix swab (all performed by Global Labs, Durban, South Africa, as previously described (Byrne et al., 2016)). Positive results were followed by a second, confirmatory assay. All women with a positive STI test were referred for further care.

Measurement of cytokines

CVL fluid samples were thawed on ice and centrifuged at 800 x g (10 min, 4°C). The supernatant was transferred to a fresh Eppendorf tube, gently mixed by pipetting and assayed immediately using a customized high-sensitivity MILLIPLEX assay (EMD Millipore) to measure concentrations of IL-1α, IL-1β, TNF-α, IFN-γ, IL-4, IL-5, IL-6, IL-10, IL-12p70, IL-13, IL-17, IL-21, IL-23, MIP-1α, MIP-1β, MIP-3α, MIG, IP-10 and I-TAC. Samples were acquired on a Bio-Plex 3D Suspension Array System (Bio-Rad), and data

were analyzed using Bio-Plex Manager (Bio-Rad). IL-8 concentrations in human CVL fluid and tissue culture supernatant, IL-6 in tissue culture supernatant and TNF-α in murine CVL fluid were measured by ELISA (eBioscience, R&D Systems). Cytokine measurements below the limit of detection were plotted as half the minimum detectable concentration for that cytokine.

Flow cytometry

Human and mouse cells were stained with a LIVE/DEAD viability dye (Invitrogen) and fluorescently labeled monoclonal antibodies specific for the following human or mouse surface markers: hCD45 (HI30), hCD3 (UCHT1), hCD4 (SK3), hCCR5 (2D7), hHLA-DR (G46-6), hCD38 (HIT2), hCD25 (2A3), hCD8 (SK1), hCD11c (B-ly6), hCD14 (M5E2) and hCD19 (HIB19), mCD3 (145-2C11) (BD), mCD45 (30-F11), mCD4 (RM4-5), mCD8a (53-6.7), and mCD44 (IM7) (BioLegend). Samples were acquired on a FACS Aria III or on an LSR II (BD). The whole human cytobrush sample and mouse cells were acquired fresh within 6 hours of collection. Mouse cells were quantified using CountBright Absolute Counting Beads (Invitrogen). Data were analyzed using FlowJo Version 9.8.5 (FlowJo Enterprise). To assess the number of activated $CD4^+$ T cells in a sample, doublets and dead cells were excluded, and the remaining events were gated for CD45, CD3 and $CD4^+$ cells and CCR5, HLA-DR and CD38 (human) or CD44 (mouse). FMO controls were used to define gates for activation markers.

Detection of Differentially Abundant Bacterial Taxa

Resolved sequence variants (RSV) were identified from 16S amplicons using the dada2 procedure (Callahan et al., 2016). Prior to RSV resolution, sequences were quality trimmed to a 220 bases. RSV were inferred from pooled sequences and taxonomy was

assigned using a naïve Bayes classifier against GreenGenes August 2013 release database (gg_13_8_99) (McDonald et al., 2012). Differentially abundant taxa were identified using the DESeq2 package (Love et al., 2014).

Bacteria

Lactobacillus crispatus (provided by Dr. David N. Fredricks) was subcultured on Columbia Blood Agar and grown in Lactobacilli MRS broth (Hardy Diagnostics) anaerobically prior to experiments. *Prevotella bivia* (ATCC 29303), *Prevotella amnii* (CCUG 53648T) and *Gardnerella vaginalis* (ATCC 14018) were cultured on Brucella Blood Agar with Hemin and Vitamin K (Hardy Diagnostics) anaerobically. Inocula were prepared in ultra-pure PBS (Invitrogen) on the day of the experiment, using a spectrophotometer (GE Healthcare) to adjust the optical density as desired, and plated out on their respective culture medium to assess the bacterial concentration.

Epithelial co-culture assays

Human endocervical epithelial cells (End1/E6E7, ATCC) were grown in keratinocyte serum-free medium (KSF) (Invitrogen) supplemented with bovine pituitary extract (0.05 mg/mL), epidermal growth factor (0.1 ng/mL), and calcium chloride (0.4 nM). For experiments, cells were plated in 96-well plates at 2.5 x 10^4 per well, and allowed to adhere for 24 hours prior to addition of bacteria. 100 ul of a 1 x 10^7 - 10^8 cfu/mL bacterial suspension in tissue culture medium was added per well and incubated anaerobically at 37 °C. After 24 h of co-culture, the supernatant was aspirated, and an aliquot from each condition plated to confirm bacterial viability. Each treatment had 4 replicates, and each experiment was repeated at least twice.

Mice

All animal procedures were approved by the Animal Resources at Children's Hospital (ARCH) review committee (protocol # 16-08-3219R). Germ-free Swiss-Webster mice were bred and housed in germ-free isolators at Children's Hospital, Boston, MA. For intravaginal inoculation experiments, 6-9 week old female mice were transferred to autoclaved cages as previously described ("out-of-the-isolator protocol") (Faith et al., 2014).

Intravaginal inoculation of mice with bacteria

Mice were anaesthetized by i.p. injection of ketamine (60 mg/kg; Santa Cruz Animal Health) and xylazine (10 mg/kg; Hospira). CVL fluid was collected by gently pipetting 20 μl of ultra-pure PBS three times up and down the vaginal cavity. This was repeated twice. Then, 20 μl of inoculum containing 2x10^8 CFU/mL were released into the vaginal cavity using a 20 μl pipette tip.

Processing of mouse tissues

Blood was collected from the facial vein, and red blood cells were removed using an RBC lysis buffer (BioLegend). Mice were sacrificed via CO_2 asphyxiation, and the mouse female reproductive tract was excised. The uterine horns were removed, and the cervix and vagina were stored in RPMI supplemented with 10% FCS at 4 °C. In the lab, the tissue was cut into tiny pieces by the use of scalpels and subjected to enzymatic digestion with 0.5 mg/mL collagenase II (Sigma) and 3 μg/mL DNase I (Qiagen) for 30 min at 37 °C. The tissue fragments were subsequently pipetted vigorously, ground on a 40 um cell strainer using the sterile stamp of a 5 mL syringe, and frequently rinsed with

ice-cold FACS buffer. The resulting cell suspensions were centrifuged and used for downstream procedures.

Statistics

Descriptive measures, such as median, interquartile range (IQR), frequency and percent, were used to summarize the data. For comparison of continuous data between two groups, Mann-Whitney test was performed. Kruskal-Wallis test with Dunn's post hoc analyses was used for comparing continuous data between more than two groups. The time to HIV acquisition was summarized using Kaplan-Meier curves, and significance was assessed by Log-rank test. Fisher's exact test was performed for comparison of categorical data between two or more groups. Principal component analysis (PCA) was used to obtain summary measures for the multi-variable cytokine set and performed using FactoMineR package in R. Multivariate analysis using MaAsLin (https://bitbucket.org/biobakery/maaslin) (Morgan et al., 2012) was performed to identify bacteria correlated with cytokine principal component 1. More specifically, linear regression of logit-transformed bacterial abundances was performed on principal component 1 scores. Bacteria not present in at least 10% of the samples with at least 0.01% abundance were excluded from the analysis. Adjustments for multiple testing were made using the Benjamini-Hochberg method.

All p values shown are two-sided. P values indicated by asterisks should be interpreted as follows: *$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<0.0001$; n.s.: $p\geq0.05$. Unless indicated otherwise, the analyses were performed in Prism 6 (GraphPad) and R Studio.

Supplemental References

Anahtar, M.N., Bowman, B.A., and Kwon, D.S. (2016). Efficient Nucleic Acid Extraction and 16S rRNA Gene Sequencing for Bacterial Community Characterization. Journal of visualized experiments : JoVE.

Anahtar, M.N., Byrne, E.H., Doherty, K.E., Bowman, B.A., Yamamoto, H.S., Soumillon, M., Padavattan, N., Ismail, N., Moodley, A., Sabatini, M.E., *et al.* (2015). Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. Immunity *42*, 965-976.

Byrne, E.H., Anahtar, M.N., Cohen, K.E., Moodley, A., Padavattan, N., Ismail, N., Bowman, B.A., Olson, G.S., Mabhula, A., Leslie, A., *et al.* (2016). Association between injectable progestin-only contraceptives and HIV acquisition and HIV target cell frequency in the female genital tract in South African women: a prospective cohort study. The Lancet Infectious Diseases *16*, 441-448.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods *13*, 581-583.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nature methods *7*, 335-336.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci Transl Med *6*.

Finkbeiner, S.R., Holtz, L.R., Jiang, Y., Rajendran, P., Franz, C.J., Zhao, G., Kang, G., and Wang, D. (2009). Human stool contains a previously unrecognized diversity of novel astroviruses. Virology journal *6*, 161.

Handley, S.A., Desai, C., Zhao, G., Droit, L., Monaco, C.L., Schroeder, A.C., Nkolola, J.P., Norman, M.E., Miller, A.D., Wang, D., *et al.* (2016). SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe *19*, 323-335.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 1-21.

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology *71*, 8228-8235.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J *6*, 610-618.

Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldridge, M.T., Wilen, C.B., Flagg, M., *et al.* (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe *19*, 311-322.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B., *et al.* (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol *13*, R79.

Oksanen, J.B., F. G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P. R.; O'Hara, R. B.; Simpson, G. L.; Solymos, P.; Stevens, M. H. H.; Szoecs, E. and Wagner, H. (2016). vegan: Community Ecology Package. R package version 2.4-0. .

Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. GigaScience *2*, 16.